DOCUMENT RESUME

ED 170 635 TM 008 998

AUTHOR Convey, John J.
TITLE Some Simultaneous Inference Procedures for A Priori Contrasts.
PUB DATE Apr 79
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Comparative Analysis; Comparative Statistics; *Hypothesis Testing; Mathematical Models; *Research Methodology; Social Science Research; *Statistical Analysis; *Statistical Studies
IDENTIFIERS A Priori Tests; Post Hoc Methods

ABSTRACT

The testing of a priori contrasts, post hoc contrasts, and experimental error rates are discussed. Methods for controlling the experimental error rate for a set of a priori contrasts tested simultaneously have been developed by Dunnett, Dunn, Sidak, and Krishnaiah. Each of these methods is discussed and contrasted as to applicability, power, and availability of tables. A strategy is suggested as to how these might be employed when a researcher desires to control the experimental error rate for a set of a priori contrasts. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED174635

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Some Simultaneous Inference Procedures For A Priori Contrasts

John J. Convey

Catholic University of America

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

John J. Convey

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Presented at the Annual Meeting of the American Educational Research Association

San Francisco, April, 1979

TM008 998

Abstract

Testing several contrasts each at a per comparison error rate of α_o allows the experimentwise error rate to exceed α_c. Methods for controlling the experimentwise error rate for a set of a priori contrasts tested simultaneously have been developed by Dunnett, Dunn, Sidak, and Krishnaiah. Each of these methods is discussed and contrasted as to applicability, power, and availability of tables. A strategy is suggested as to how these might be employed when a researcher desires to control the experimentwise error rate for a set of a priori contrasts.

Some Simultaneous Inference Procedures for a Priori Contrasts

Researchers in the social sciences often are concerned with questions involving the comparisons of means from several groups. In many cases, an omnibus hypothesis of the form: $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$, is tested and, if rejected, is followed by a series of contrasts to determine whether various combinations of means are significantly different from each other or from zero. Such comparisons are post-hoc since they follow the testing of an omnibus hypothesis. In other cases, the omnibus hypothesis is not tested since the researcher prior to observing the data specifies which combinations of means are of interest and directly tests each. These comparisons are a priori in nature.

In each of the above cases, typically more than one comparison is of interest. Thus, the problem of controlling Type I error becomes an issue. At least two Type I error rates exist for a given set of comparisons (Miller, 1966; Petrinovich & Hardyck, 1969): a per comparison error rate and an experimentwise error rate. The probability of a Type I error for any given contrast is the per comparison error rate, while the probability of a Type I error for several contrasts considered simultaneously is the experimentwise error rate.

It is known that testing several contrasts each at a per comparison error rate of α_0 allows the experimentwise error rate to exceed α_0.

The degree of inflation of the experimentwise error rate over the per comparison error rate depends on the number of contrasts and the relationship between them. For a fixed per comparison error rate, the experimentwise error rate increases as the number of contrasts increases and/or as the correlation between the contrasts decreases. For example, if p contrasts are tested simultaneously and the contrasts are independent (i.e., $\rho^1_{ij} = 0$, $i \neq j = 1, 2, \ldots, p$), then the experimentwise error rate α_{exp} can be approximated by:

$$\alpha_{exp} = 1 - (1 - \alpha_1)(1 - \alpha_2) \ldots (1 - \alpha_p),$$

where α_i is the per comparison error rate for the ith contrast. Hays (1972) points out that this is an approximation, since usually a common error term is used to test each contrast so that the tests are not independent. If five independent contrasts are each tested at a per comparison error rate of .05, then the approximate experimentwise error rate is $1 - (1 - .05)^5 = .226$. Because of the potentially high probability of a Type I error for an experiment, the experimentwise error rate should not be overlooked when multiple contrasts are considered (Games, 1971; Miller, 1966; Tukey, Note 1).

Depending on the nature of the set of contrasts and on the relative seriousness of inflating the experimentwise error rate, different procedures have been developed for dealing with multiple comparisons. Some procedures are designed for dealing with comparisons which are post hoc, while others are more suitable for a priori comparisons. Generally, the procedures for dealing with post hoc comparisons are well-known to researchers in the social sciences. The major post hoc techniques include Tukey's Wholly Significant

Difference Test or the T-Method (Tukey, Note 1), Scheffé's F-projections or the S-Method (Scheffé, 1953), Fisher's Least Significant Difference Test (Fisher, 1935), Newman-Keuls Multiple Range Test (Newman, 1939; Keuls, 1952), and Duncan's New Multiple Range Test (Duncan, 1955, 1957). Comprehensive reviews of these procedures have been provided by Games (1971), Miller (1966) and O'Neil and Wetherhill (1971). These procedures differ in the control they provide both for the experimentwise error rate and the per comparison error rate. Of these, only the T-Method and the S-Method control the experimentwise error rate precisely at a specified α_{exp}. The T-Method controls the experimentwise error rate when all possible pairwise contrasts among k groups of equal size are considered. (Some extensions of the T-Method have been reviewed by Keselman and Rogan, 1977.) The S-Method controls the experimentwise error rate when all possible linear contrasts among k means are considered. Generally, those procedures that control precisely the experimentwise error rate, do so at the expense of power (Games, 1971).

When the researcher is able to specify the contrasts of interest a priori and desires to maintain the experimentwise error rate below a certain level, the use of a post hoc technique often results in considerable loss of power. In most cases, the a priori contrasts do not include all possible pairwise contrasts nor a large percentage of all possible linear contrasts. Thus, the use of the T-Method or the S-Method in these cases would not be very efficient in terms of power. Several methods are available to control the experimentwise error rate for a priori contrasts. Two of these, Dunnett's Many-One

t-statistics (Dunnett, 1955; 1964) and Bonferroni t-statistics (Dunn, 1961) frequently have been used in social science research. More recently, Šidak (1968, 1971) and Krishnaiah (1965a, 1965b, 1965c, 1969, Note 2, Note 3) have developed methods which are more widely applicable than Dunnett's t-statistics and more powerful than Bonferroni t-statistics. The purpose of this paper is to provide a brief overview of these four methods, contrast them as to power and availability, and to suggest a strategy that a researcher might employ when confronted with a set of a priori contrasts and a desire to control the experimentwise error rate.

Dunnett's Many-One t-statistics

Often the researcher is interested in comparing each of $k - 1$ treatment means against the mean of a control group. Since $k - 1$ contrasts are being considered simultaneously, the researcher may want to control the experimentwise error rate at α_{exp} for all $k - 1$ contrasts. Dunnett (1955) proposed a procedure to control the experimentwise error rate at α_{exp} in this situation.

Let $H_{ik}: \mu_i - \mu_k$, where $\mu_i = 1, 2, \ldots, k - 1$ represents the means of the treatment groups and μ_k represents the mean of the control group. A t-statistic is calculated as usual for each contrast:

$$t_{ik} = \frac{(\bar{X}_i - \bar{X}_k)}{\sqrt{MS_{error}(\frac{1}{n_i} + \frac{1}{n_j})}}, \quad i = 1, 2, \ldots, k - 1.$$

When $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ is true, the joint distribution of $t_{1k}, t_{2k}, \ldots, t_{k-1,k}$ is a multivariate t distribution with $n - k$

degrees of freedom and with $\Omega = (\rho_{ij})$ as the correlation matrix of the accompanying multivariate normal distribution. For the many-one case, Ω is a matrix of 1's on the diagonal and .5's on the off-diagonals. The multivariate t distribution was first considered by Dunnett and Sobel (1954, 1955) and Cornish (1954), independently. Critical values for the Dunnett statistic can be found in Dunnett (1964) and in many testbooks on design (see, for example, Winer, 1971).

Use of the multivariate t distribution controls for the experimentwise error rate in this situation. For example, suppose $k = 6$, $\alpha = .05$, $df = 20$, and two-tailed alternatives are desired. The critical value of t if each comparison is considered separately would be 2.09, while the Dunnett critical value would be 2.65. The Dunnett critical value in this case satisfies:

$$P(|t_{ik}| < t_{\alpha};\ i = 1, \ldots, 5 | \mu_1 = \ldots = \mu_5) = .95$$

Dunnett's method then is applicable only to a special set of a priori contrasts; viz., those in which the correlation between any two contrasts is .5. This situation occurs when the $k - 1$ treatment means are compared to a control mean. In this special situation, Dunnett's test provides an exact control of the experimentwise error rate at some prespecified values. The test is applicable for both one and two-tailed alternatives.

Bonferroni t-Statistics

Dunn (1961), using the Bonferroni or Boole inequality, showed that the experimentwise error rate α_{exp} cannot exceed the sum of the per comparison error rates for each contrast (or linear combination) being

tested. That is,

$$\alpha_{exp} < \sum_{i=1}^{p} \alpha_i \qquad (1)$$

where, α_i is the per comparison error rate for contrast i. Thus, if each of p contrasts are tested at the $\alpha/2p$ level of significance against two-tailed alternatives (α/p for one-tailed alternatives), then the experimentwise error rate cannot exceed α.

Dunn (1961) and Miller (1966) show that the Bonferroni inequality is applicable in many situations and not limited to sets of contrasts on means. Given a set of p contrasts and a desire to have the experimentwise error rate to be at most .05, the researcher simply calculates a t statistic, t_i, for each contrast c_i, i = , 2, ..., p, and compares each t_i to a critical value t_{α_i} obtained such that the sum of the α_i, i = 1, ..., p does not exceed .05. The procedure is flexible in that it is not necessary for $\alpha_i = \alpha_j$, $i \neq j$; only that $\sum_{i=1}^{p} \alpha_i < .05$. So if $\alpha_{exp} = .05$ and $p = 5$, each contrast could be tested at $\alpha_i = .05/5 = .01$, i = 1, ..., 5; or, if there were reason to do so, the following per comparison error rates could be used: $\alpha_1 = .03, \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = .005$. Dunn (1961) gives tables of critical values of $t_{\alpha/2p}$ for select values of df and number of comparisons.

Because of the strict inequality of Equation 1, only an upper bound for the experimentwise error rate is obtained; thus the procedure is less powerful than one for which the exact experimentwise error rate is known. Due to the nature of the Bonferroni inequality, the difference

in power between the Bonferroni procedure and an exact procedure increases as the number of contrasts increases. In addition, the difference in power is a function of the correlation between the contrasts, the degrees of freedom associated with the standard error, and the experimentwise error rate. These relationships will be explored later in this paper. Despite the difference in power between the Bonferroni procedure and exact procedures, Dunn (1961) and Miller (1966, 1977) were impressed by the simplicity and competitiveness of the Bonferroni procedure. As will be seen, generally, the loss in power from using the Bonferroni procedure over an exact procedure is slight.

The exact procedures that provide competitors to the Bonferroni procedure are the Finite Intersection Tests of Krishnaiah (1965a, Note 2, Note 3). In addition, a special case of the Finite Intersection Tests developed by Sidak provides a more powerful test than Bonferroni. These procedures will be discussed next.

Sidak's Procedure

Let $\underline{x}' = (\underline{x}_1, \ldots, \underline{x}_p)$ have a multivariate normal distribution with zero mean vector and arbitrary covariance matrix. Let $\underline{S}$ be an independent random variable distributed as $\chi^2_{\underline{df}}/\underline{df}$. Sidak (1967, 1971) proved that

$$P\left(\frac{|x_1|}{\underline{S}} \leq \underline{c}_1, \ldots, \frac{|x_p|}{\underline{S}} \leq \underline{c}_p\right) \geq \prod_{\underline{i}=1}^{p} P\left(\frac{|x_i|}{\underline{S}} \leq \underline{c}_i\right) \qquad (2)$$

That is, the probability that a random normal vector with arbitrary covariances falls inside a rectangle centered at its means is always

at least as large as the corresponding probability in the case where the variables are independent. Equation 2 can be used to show that

$$1 - \alpha_{exp} \geq (1 - \alpha)^{p} > 1 - p\alpha \tag{3}$$

or

$$\alpha_{exp} \leq 1 - (1 - \alpha)^{p} < p\alpha, \tag{4}$$

where α is the per comparison error rate and p is the number of comparisons.

For fixed values of p and α_{exp}, the critical value resulting from the second term of Equations 3 or 4 will be less than the critical value using the third term. Thus, the use of the Sidak result would produce slightly smaller intervals or slightly more powerful tests than the Bonferroni procedure. If orthogonal contrasts are used, Equation 2 is strictly an equation and the Sidak procedure is exact. For other than orthogonal contrasts, Sidak's procedure is approximate, but always more powerful than Bonferroni due to Equations 3 and 4. Critical values for the Sidak procedure can be obtained from tables provided by Armitage and Krishnaiah (Note 4, Note 5) and from charts provided by Moses (Note 6).

Finite Intersection Tests

Krishnaiah (1965a) proposed general procedures for testing p hypothese H_{01}, ..., H_{0p} against appropriate alternatives. These hypotheses could be multivariate or univariate, and need not be limited to contrasts. The procedures are called Finite Intersection Tests, since they form the intersection of the hypotheses of interest and test these simultaneously using the joint distribution of the set of

test statistics appropriate for testing each hypothesis separately. In so doing, these procedures maintain the experimentwise error rate at a specified value, α_{exp}. Complete descriptions of these procedures are found in Krishnaiah (1965a, 1965b, 1965c, 1969, Note 2, Note 3).

In this section, only the situation in which all hypotheses are univariate and consist solely of contrasts will be discussed. Suppose a series of p hypotheses $H_{01}, \ldots, H_{0p}$ are to be tested against two-tailed alternatives. The joint test of $H_{01}, \ldots, H_{0p}$ involves the joint distribution of $F_1, \ldots, F_p$ where F_i is the square of the ratio of the ith contrast and the standard error of the ith contrast; i.e., $F_i = (c_i/s_{c_i})^2$. That is, when $\bigcap_{i=1}^{p} H_{0i}$ is true, $F_1, \ldots, F_p$ is distributed jointly as a p-variate F distribution with parameters 1, df and with Ω as the correlation matrix of the accompanying multivariate normal distribution. Any H_{0i} is rejected when $F_i > W_\alpha$, where W_α is chosen so that:

$$P(F_i < W_\alpha,\ i = 1, 2, \ldots, p \mid \bigcap_{i=1}^{p} H_{0i}) = 1 - \alpha.$$

If the alternative hypotheses are one-tailed, then the joint test of $H_{01}, \ldots, H_{0p}$ involves the joint distribution of $t_1, \ldots, t_p$ or a p-variate t distribution. Any H_{0i} is rejected when $t_i > t_\alpha$, where t_α is chosen so that:

$$P(t_i < t_\alpha,\ i = 1, 2, \ldots, p \mid \bigcap_{i=1}^{p} H_{0i}) = 1 - \alpha.$$

The multivariate F distribution considered here was presented by Krishnaiah (1965a, 1965c) and is a generalization of the multivariate

t distribution presented by Dunnett and Sobel (1954, 1955) and Cornish (1954). Because the probability density function for the multivariate F and multivariate t distributions depend upon the number of contrasts and the correlations between them, compilation of exact percentage points for these distributions are available only when Ω is non-singular and for equicorrelated contrasts; i.e., $\rho_{ij} = \rho(i \neq j = 1, 2, \ldots, p)$. For testing two-tailed alternatives, upper 5% and 1% points of the distribution of the maximum t^2 variates which are jointly distributed as the p-variate F are tabulated by Krishnaiah and Armitage (1970) for $df = 5(1)35$. Dunn and Massey (1965) present tables of the distribution of $\text{Max}_i|t_i|$ for $1 - \alpha = .50(.10).90, .95, .975, .99$; $\rho = .0(.1)1.0$, $p = 1, 6, 10, 20$ and $df = 4, 10, 30, \infty$. These tables can be used to obtain critical values for the multivariate F. For testing one-tailed alternatives, upper 5% and 1% points of the multivariate t distribution are given by Krishnaiah and Armitage (1966) for $df = 5(1)35$.

It should be noted that in the equicorrelated situation, Dunnett's procedure and Sidak's procedure are special cases of the Finite Intersection Tests. When $\rho = .5$, the Finite Intersection Tests reduce to Dunnett's procedure; and when $\rho = 0$, to Sidak's procedure.

When Ω is singular or when not all contrasts are equicorrelated, exact percentage points for the F distributions are not available. In these situations, Krishnaiah (1965a, Note 2, Note 3) used Poincare's formula to place monotonic upper and lower bounds on $P(F_i < W_\alpha)$ and $P(t_i < t_\alpha)$. From these bounds, upper and lower bounds on the critical values can be obtained. The upper bound for the critical value is the value provided by Sidak's procedure; the calculation of the lower

bound is more problematic. Krishnaiah using Poincare's formula suggested the following approximation:

$$P(\underline{F}_i < \underline{W}_\alpha;\ i = 1, 2, \ldots, p) \leq 1 - P_0 + P_1 \tag{5}$$

where,
$$P = \sum_{i=1}^{p} P(\underline{F}_i \geq \underline{W}_\alpha) \tag{6}$$

and
$$P = \sum_{i<j} P(\underline{F}_i \geq \underline{W}_\alpha,\ \underline{F}_j \geq \underline{W}_\alpha) \tag{7}$$

Probability values for Equation 6 can be read from ordinary central $\underline{F}$ tables. Probability values for Equation 7 are provided by Krishnaiah, Armitage, and Breiter (Note 7, Note 8). However, what is needed is the critical value and this cannot be obtained easily from an inspection of these tables without the help of a computer. A computer program has been developed to estimate the critical values for this inequality (Krishnaiah, Note 3). Since the probability statement in Equation 5 provides an upper bound, the corresponding critical value will be a lower bound for the exact critical value.

Comparison

Under the conditions of a non-singular Ω and equicorrelated contrasts, the Finite Intersection Tests are exact procedures in that the experimentwise error rate is known and is as specified. On the other hand, the Bonferroni procedure is approximate in that the experimentwise error rate is not known, but is less than the specified value. Under these conditions then, the Finite Intersection Tests are more powerful than the Bonferroni procedure. Even under the condition of a singular Ω or nonequicorrelated contrasts, the Finite

Intersection Tests are more powerful than Bonferroni, since Sidak's procedure produces shorter intervals than Bonferroni and also provides an upper bound for the critical value of the Finite Intersection Tests. Theoretically, then, the researcher should not use the Bonferroni procedure to control the experimentwise error rate since a more powerful alternative always exists.

In practice, however, the difference in power between the Bonferroni procedure and the Finite Intersection Tests, including the special cases of Sidak and Dunnett, usually is slight. As noted previously, the difference in power between the Bonferroni procedure and an exact procedure is a function of the number of contrasts (p), the correlation between the contrasts (ρ), the experimentwise error rate (α_{exp}), and the degrees of freedom associated with the standard error of the contrast (df). For an experimentwise error rate of .05 and .01 respectively, Table 1 and Table 2 give the critical values of t for the Bonferroni procedure and the Finite Intersection Tests in the equicorrelated case ($\rho = 0$, $\rho = .5$, and $\rho = .8$) for $p = 2(1)10$ and $df = 15$ and 30. The Bonferroni critical values in these tables were obtained from tables provided by Dunn (1961); the values for $\rho = 0$, from tables provided by Armitage and Krishnaiah (Note 4); and the values for $\rho = .5$ and $\rho = .8$, from tables provided by Krishnaiah and Armitage (1970).

Insert Table 1 and Table 2 about here

In order to compare the relative power of the exact procedures at $\rho = 0$, $\rho = .5$, and $\rho = .8$ with the Bonferroni procedure, an efficiency ratio was constructed by dividing the critical value of each of these by the Bonferroni critical value. Table 3 shows these efficiency ratios for two, five, and ten contrasts at $\underline{df} = 15$ and 30, and with an experimentwise error rate of .05 and .01. An efficiency ratio $\underline{r}_0$ indicates that the length of the confidence intervals constructed using that critical value will be $\underline{r}_0$ times the length of the confidence intervals constructed by the Bonferroni critical value. If $\underline{r}_0 < 1$, then the intervals are shorter than the Bonferroni intervals and hence more powerful. For example, when $\rho = .8$, $\alpha_{exp} = .05$, and $\underline{p} = 10$, then $\underline{r}_0 = .854$. This indicates that the length of the intervals constructed by the Finite Intersection Tests under these conditions will be .854 times the length of the Bonferroni intervals or 14.6% shorter. Since shorter intervals indicate more powerful procedures, provided that other parameters are equal, the Finite Intersection Tests are seen to be more powerful than the Bonferroni procedure.

Insert Table 3 about here

The examples in Table 3 clearly show the difference in power between the exact procedures and the Bonferroni procedure as a function of number of contrasts, correlation between the contrasts, degrees of freedom for the standard error, and experimentwise error rate. For

a fixed α_{exp} and $\underline{df}$, $\underline{r}_0$ decreases when $\underline{p}$ increases or when ρ increases. Thus, the use of the Finite Intersection Tests becomes more desirable the larger the number of contrasts to be tested simultaneously or the larger the correlation between the contrasts. When $\underline{p}$ and ρ are fixed, $\underline{r}_0$ increases as $\underline{df}$ increases or as α_{exp} decreases. One important consequence of this is that little difference exists between the Bonferroni procedure and the Finite Intersection Tests when the sample sizes are large.

Generally, the values of $\underline{r}_0$ are close to 1, indicating only a slight advantage in favor of the Finite Intersection Tests. In the case of Sidak's procedure ($\rho = 0$), the greatest advantage depicted in Table 3 occurs when $\underline{p} = 10$, $\underline{df} = 15$, and $\alpha_{exp} = .05$. In this case, the Sidak intervals are 1.8% shorter than the Bonferroni intervals. In most situations this difference hardly will make any difference in the researcher's statistical decisions. Thus, when orthogonal contrasts are being considered, even though Sidak provides an exact test, in most situations it appears that little will be lost using the Bonferroni procedure. When the sample sizes are small, differences in statistical decisions may tend to occur as the correlations increase and as the number of contrasts increase.

In most practical situations, due either to the nature of the contrasts that are of interest or to unequal group sizes, the contrasts are not equicorrelated. In these situations, the use of the Finite Intersection Tests involves the calculation of upper and lower bounds on the critical value. We have seen that Sidak's procedure provides the upper bound while the lower bound has to be estimated from

Poincare's inequality. An example of the power efficiency of the Finite Intersection Tests in a case like this is given below. This example is adapted from one provided by Krishnaiah (Note 9).

Consider a situation where samples of size 9, 7, 8, and 10 are drawn from four populations, respectively. Four sets of contrasts are of interest with the experimentwise error rate set at .05 for each set. These sets are as follows:

Set 1: all pairwise contrasts ($p = 6$)

e.g. $\mu_1 - \mu_2 = 0$; $\mu_1 - \mu_3 = 0$; etc.

Set 2: the average of any three groups against the remaining group ($p = 4$)

e.g. $\mu_1 + \mu_2 + \mu_3 - 3\mu_4 = 0$; etc.

Set 3: adjacent pairwise contrasts ($p = 3$)

e.g. $\mu_1 - \mu_2 = 0$; $\mu_2 - \mu_3 = 0$; $\mu_3 - \mu_4 = 0$.

Set 4: each of the first three groups against the fourth ($p = 3$)

e.g. $\mu_1 - \mu_4 = 0$; $\mu_2 - \mu_4 = 0$; $\mu_3 - \mu_4 = 0$.

In none of these sets are the contrasts equicorrelated. Table 4 shows the critical values of the Bonferroni procedure, and the upper and lower critical values and efficiency ratios of the Finite Intersection Tests at $df = 30$ and $p = 3$, 4, and 6. It is apparent from Table 4 that, in this isolated example, the intervals produced by the upper and lower critical values of the Finite Intersection Tests do not improve substantially upon the intervals provided by the Bonferroni procedure.

Insert Table 4 about here

In some instances when the contrasts are not equicorrelated it may be possible to improve slightly upon the Sidak upper bound by dividing the contrast set into homogeneous equicorrelated subsets, thus providing an exact test for the contrasts in these subsets. If the researcher wished to control the experimentwise error rate at α_{exp} for all the contrasts, then each subset of contrasts can be tested at $\alpha_{exp/g}$, where g is the number of subsets. For example, consider a situation with six groups and six observations per group. The contrast set of interest is: $\mu_1 - \mu_2 = 0$, $\mu_1 - \mu_3 = 0$, $\mu_4 - \mu_6 = 0$, and $\mu_5 - \mu_6 = 0$. The Bonferroni value at $df = 30$ and $\alpha_{exp} = .05$ is 2.66, while the Sidak value is 2.64. If two homogeneous subsets are formed as follows:

$$\begin{cases} \mu_1 - \mu_2 = 0 \\ \mu_1 - \mu_3 = 0 \end{cases} \qquad \begin{cases} \mu_4 - \mu_6 = 0 \\ \mu_5 - \mu_6 = 0 \end{cases}$$

and each is tested at $df = 30$ and $.05/2 = .025$, the critical value can be improved to 2.63 using Dunnett's test, since the correlation in the subsets is .5. In this example, the improvement over the Sidak bound and the Bonferroni procedure is slight. Generally, it appears that any improvement provided by this technique will be only slight compared with the use of the Sidak or Bonferroni procedures. Furthermore, in many cases precise critical values will be unavailable due to the lack of appropriately detailed tables.

Conclusions and Recommendations

The Finite Intersection Tests are exact in the equicorrelated case and produce upper and lower bounds to the exact critical value in the nonequicorrelated case. While the Finite Intersection Tests

always will be more powerful than the Bonferroni procedure, the difference in power will usually be slight. The difference in power between these procedures increases as the number of comparisons increases, the correlations between the comparisons increase, and/or the group sizes decrease. The Bonferroni procedure is very easy to apply and is quite flexible, allowing different per comparison rates for different contrasts within a set while controlling the experiment-wise error for the set. In the equicorrelated case, the Finite Intersection Tests are easy to use and tables are readily available. However, when the correlations between the contrasts are not all the same, the use of the Finite Intersection Tests becomes somewhat cumbersome.

Sidak's procedure is applicable to all sets of contrasts and is exact for orthogonal contrasts. In the nonequicorrelated case, Sidak's procedure produces the upper bound for the critical value of the Finite Intersection Tests. For any contrast set, the Sidak critical value will be less than the Bonferroni critical value, however the difference in power between these procedures is very slight. Bonferroni's procedure is more flexible than Sidak's in that the per comparison error rate for Sidak's is the same for each contrast, while the per comparison error rate need not be uniform in the Bonferroni procedure. When the correlations between the contrasts are each .5, Dunnett's test is more powerful and tables of critical values are readily available. Both Sidak's procedure and Dunnett's procedure are special cases of the Finite Intersection Tests.

When the researcher has specified a number of a priori contrasts and wishes to control the experimentwise error rate for these contrasts, it is recommended that the Bonferroni strategy be applied routinely. In most cases when a true difference exists among the means at the specified probability level, the Bonferroni procedure will be able to detect it. In the event that an observed value for any contrast in the set is slightly less than the critical value provided by the Bonferroni strategy, the use of the Sidak or the Finite Intersection Tests may provide a significant finding.

In the event the contrasts of interest are orthogonal, the Sidak procedure should be applied since it is exact in this case and relatively straightforward. Dunnett's test should be used when the correlations between the contrasts are all equal to .5 since it is exact in that case.

Finally, if the researcher is interested in constructing confidence intervals rather than in testing hypotheses, use of the Finite Intersection Tests always will produce tighter intervals than those produced by the Bonferroni procedure. However, since the critical values for the Bonferroni procedure are usually very close to the critical values for the Finite Intersection Tests, use of the Bonferroni procedure will produce good approximations to those intervals produced by the Finite Intersection Tests. Theoretically, exact procedures are preferable to approximate procedures. However, in the case of a priori contrasts, little is sacrificed by using the approximate Bonferroni procedure.

Reference Notes

1. Tukey, J. W. The problem of multiple comparisons. Unpublished manuscript, Princeton University, 1953.

2. Krishnaiah, P. R. Simultaneous inference procedures. Paper presented at the annual meeting of the American Educational Research Association, New York, April, 1977.

3. Krishnaiah, P. R. Some developments on simultaneous test procedures. Unpublished manuscript, University of Pittsburgh, Department of Mathematics and Statistics, 1978.

4. Armitage, J. V. and Krishnaiah, P. R. Tables for the studentized largest chi-square distribution and their application (ARL 64-188). Wright-Patterson Air Force Base, Ohio, 1964.

5. Armitage, J. V. & Krishnaiah, P. R. Approximate percentage points for applications of finite intersections tests for multiple comparisons of means. Unpublished manuscript, 1978. (Available from Department of Mathematics and Statistics, University of Pittsburgh.)

6. Moses, L. E. Charts for finding upper percentage points of Student's t in the range of .01 to .00001 (Technical report No. 24). Stanford University, 1976.

7. Krishnaiah, P. R., Armitage, J. V., & Breiter, M. C. Tables for the probability integrals of the bivariate t distribution (ARL 69-060). Wright-Patterson Air Force Base, Ohio, 1969.

8. Krishnaiah, P. R., Armitage, J. V., & Breiter, M. C. Tables for the bivariate t distribution (ARL 69-210). Wright-Patterson Air Force Base, Ohio, 1969.

9. Krishnaiah, P. R. Computer printout of an example presented at the Short Course on Multivariate Statistics prior to the Fifth International Symposium on Multivariate Statistics, University of Pittsburgh, June 18, 1978.

References

Cornish, E. A. The multivariate small t-distribution associated with a set of normal sample deviates. Australian Journal of Psychology, 1954, 7, 531-542.

Duncan, D. B. Multiple range and multiple F tests. Biometrics, 1955, 11, 1-42.

Duncan, D. B. Multiple range tests for correlated and heteroscedastic means. Biometrics, 1957, 13, 164-176.

Dunn, O. J. Multiple comparisons among means. Journal of the American Statistical Association, 1961, 56, 52-64.

Dunn, O. J. & Massey, F. J. Estimation of multiple contrasts using t-distributions. Journal of the American Statistical Association, 1965, 60, 573-583.

Dunnett, C. W. New tables for multiple comparisons with a control. Biometrics, 1964, 20, 482-491.

Dunnett, C. W. A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association, 1955, 50, 1096-1121.

Dunnett, C. W. & Sobel, M. A bivariate generalization of Student's t-distribution, with tables for certain cases. Biometrika, 1954, 41, 153-169.

Dunnett, C. W. & Sobel, M. Approximations to the probability integral and certain percentage points of a multivariate analogue of student's t-distribution. Biometrika, 1955, 42, 258-260.

Fisher, R. A. The design of experiments. Edinburgh & London: Oliver & Boyd, Ltd., 1935.

Games, P. A. Multiple comparisons of means. American Educational Research Journal, 1971, 8, 531-565.

Hays, W. L. Statistics for the social sciences (2nd ed.) New York: Holt, Rinehart & Winston, 1972.

Keselman, H. J. & Rogan, J. C. The Tukey multiple comparison test 1953-1976. Psychological Bulletin. 1977, 84, 1050-1056.

Keuls, M. The use of the "studentized range" in connection with an analysis of variance. Euphytica, 1952, 1, 112-122.

Krishnaiah, P. R. On the simultaneous ANOVA and MANOVA tests. Annals of the Institute of Statistical Mathematics, 1965, 17, 35-53. (a)

Krishnaiah, P. R. Multiple comparison tests in multiresponse experiments. Sankhya, Series A. 1965, 27, 65-72. (b)

Krishnaiah, P. R. On a multivariate generalization of the simultaneous analysis of variance test. Annals of the Institute of Statistical Mathematics, 1965, 17, 167-173. (c)

Krishnaiah, P. R. Simultaneous test procedures under general MANOVA models. In P. R. Krishnaiah (Ed.) Multivariate analysis - II. New York: Academic Press, 1969.

Krishnaiah, P. R. & Armitage, J. V. Tables for a multivariate t distribution. Sankhya, Series B, 1966, 28, 31-56.

Krishnaiah, P. R. & Armitage, J. V. On a multivariate F distribution. In R. C. Bose, I. M. Chakravorti, P. C. Mahalanobis, C. R. Rao, & K. J. C. Smith (Eds.) Essays in probability and statistics. Chapel Hill: University of North Carolina Press, 1970.

Miller, R. G. Simultaneous statistical inference. New York: McGraw-Hill, 1966.

Miller, R. G. Developments in multiple comparisons: 1966-1976. Journal of the American Statistical Association, 1977, 72, 779-788.

Newman, D. The distribution of range in samples from a normal population, expressed in terms of independent estimates of a standard deviation. Biometrika, 1939, 31, 20-30.

O'Neil, R. & Wetherhill, G. B. The present state of multiple comparison methods (with discussion). Journal of the Royal Statistical Society, Series B, 1971, 33, 218-250.

Petrinovich, L. F. & Hardyck, C. D. Error rates for multiple comparison methods: Some evidence concerning the frequency of erroneous conclusions. Psychological Bulletin, 1969, 71, 43-54.

Scheffé, H. A method for judging all contrasts in the analysis of variance. Biometrika, 1953, 40, 87-104.

Schuurmann, F. J., Krishnaiah, P. R., & Chattopadhyay, A. K. Tables for a multivariate F distribution. Sankhya, Series B, 1975, 37, 308-331.

Sidák, Z. Rectangular confidence regions for the means of multivariate normal distributions. Journal of the American Statistical Association, 1967, 62, 626-633.

Sidak, Z. On probabilities of rectangles in multivariate student distribution: Their dependence on correlations. Annals of Mathematical Statistics, 1971, 42, 169-175.

Winer, B. J. Statistical principles in experimental design. Second edition. New York: McGraw-Hill, 1971.

Footnote

[1]Let $C_1, C_2, \ldots, C_p$ represent a set of contrasts among k means such that:

$$C_i = a_{i1}\mu_1 + \ldots + a_{ik}\mu_k.$$

Then the correlation ρ_{ij} between two contrasts C_i and C_j is defined as:

$$\rho_{ij} = \frac{\sum_{r=1}^{k} \frac{a_{ir}a_{jr}}{n_r}}{\left(\sum_{r=1}^{k} \frac{a_{ir}^2}{n_r}\right)^{\frac{1}{2}} \left(\sum_{r=1}^{k} \frac{a_{jr}^2}{n_r}\right)^{\frac{1}{2}}}$$

Table 1

Critical Values of t as a function of df, number of contrasts, and correlation between the contrasts for an Experimentwise Error Rate of .05

p =	2	3	4	5	6	7	8	9	10
					df = 15				
Bonferroni	2.49	2.69	2.84	2.95	3.04	3.11	3.18	3.24	3.29
ρ = .0	2.47	2.67	2.81	2.91	2.99	3.06	3.13	3.18	3.23
ρ = .5	2.44	2.61	2.73	2.82	2.89	2.95	3.00	3.04	3.08
ρ = .8	2.36	2.48	2.57	2.63	2.68	2.72	2.75	2.78	2.81
					df = 30				
Bonferroni	2.36	2.54	2.66	2.75	2.83	2.89	2.94	2.99	3.03
ρ = .0	2.35	2.52	2.64	2.73	2.81	2.87	2.92	2.96	3.00
ρ = .5	2.32	2.47	2.58	2.66	2.72	2.77	2.82	2.85	2.89
ρ = .8	2.25	2.36	2.44	2.50	2.54	2.57	2.61	2.63	2.66

Table 2

Critical Values of t for an Experimentwise Error Rate of .01 for the Bonferroni Procedure and the Finite Intersection Tests for select values of df, number of contrasts and correlations.

p =	2	3	4	5	6	7	8	9	10
					df = 15				
Bonferroni	3.29	3.48	3.62	3.74	3.82	3.90	3.97	4.02	4.07
ρ = .0	3.28	3.47	3.61	3.72	3.80	3.87	3.94	3.99	4.04
ρ = .5	3.25	3.44	3.55	3.64	3.71	3.77	3.83	3.88	3.92
ρ = .8	3.18	3.31	3.40	3.47	3.52	3.56	3.60	3.63	3.66
					df = 30				
Bonferroni	3.03	3.19	3.30	3.39	3.46	3.52	3.57	3.61	3.65
ρ = .0	3.03	3.18	3.29	3.38	3.45	3.50	3.55	3.60	3.64
ρ = .0	3.01	3.15	3.25	3.33	3.39	3.44	3.49	3.52	3.56
ρ = .0	2.95	3.07	3.14	3.20	3.24	3.28	3.31	3.34	3.36

Table 3

Efficiency Ratios of the Finite Intersection Tests to the Bonferroni Procedure for Select Values of df, p, ρ, and α_{exp}

p =	2	5	10
	α_{exp} = .05; df = 15		
ρ = .0	.992	.986	.982
ρ = .5	.980	.956	.936
ρ = .8	.948	.892	.854
	α_{exp} = .05; df = 30		
ρ = .0	.996	.993	.990
ρ = .5	.983	.967	.954
ρ = .8	.953	.909	.878
	α_{exp} = .01; df = 15		
ρ = .0	.997	.995	.993
ρ = .5	.988	.973	.963
ρ = .8	.967	.928	.899
	α_{exp} = .01; df = 30		
ρ = .0	.999	.997	.997
ρ = .5	.993	.982	.975
ρ = .8	.974	.944	.921

Table 4

Bonferroni Critical Values and Finite Intersection Tests

Upper and Lower Critical Values and Efficiency Ratios

For α_{exp} = .05, df = 30, and select values of p.

p =	3	4	6
		Critical Values	
Bonferroni	2.54	2.66	2.83
Upper Bound	2.52	2.64	2.81
Lower Bound	2.48	2.61	2.68
		Efficiency Ratio	
Upper Bound	.992	.992	.991
Lower Bound	.976	.981	.947